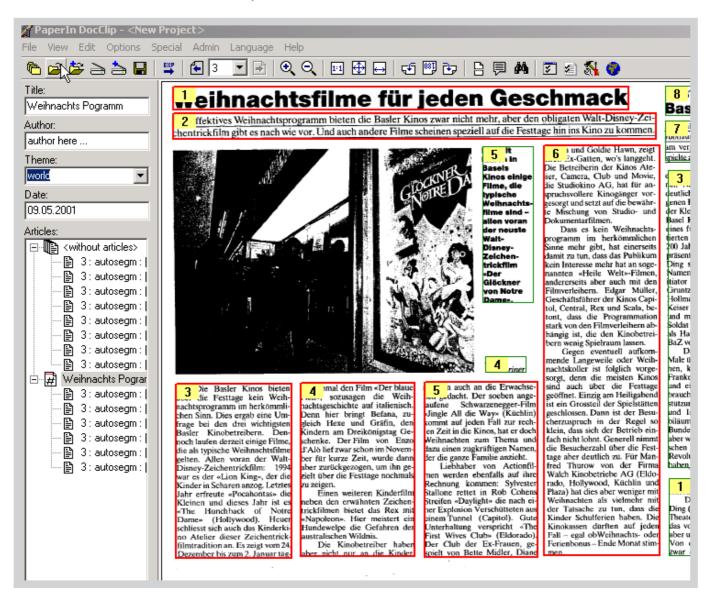


DocClip

Creating an archive with news papers and magazines, books and other documents

PaperIn DocClip with OCR (Optical Character Recogonition) allows searching in full text retrieval databases as well as sql based databases



PaperIn DocClip...

scans documents, assembles articles which are referenced tet related to the image.

segments and classifies pages in text and image parts.

allows the additional definition of index fields.

 allows the recognition of text (OCR) from chosen image segments.

 allows the recognition of most West European languages as well as Cyrillic. allows the validation of the recognized text and its correction with help of special dictionaries.

creates text files in different output formats where some of them allows a visual referencing of the recognized text on the scanned image.

writes a protocol for controlling purposes with users, date, time.

 allows the export of text files for loading directly into databases.

PaperIn DocClip...

Basic module: Scaning, segmenting, OCR, structured text output (ASCII, XML ...)

Optionale modules:

- PDF output
- Validation of text with help of language lexica (CIS)
- Linguistic analysis to distinguish between different meanings *
- Redlining (ancor and hyperlink editor)*
- Gothic font recognition *
- Arabic screen layout and segmentation *
- * in work

Technical specifications:

Character recognition/fonts: printed fonts (Omnifont) Character size: 8-72 dots at minimum 200 dpi Area: Segmented fields: graphical menu driven screen layout Image saving: Various resolutions and formats, e.g. CCITT G3/G4 Data output for text: ASCII, ISO, Xdoc², PDF³, XML For databases: Reference of image to outside area of database or embedded in file. Image acceptance: different resolutions and bilevel formats, e.g. CCITT G3/G4, GIF, JPEG Operating system: NT 4.SP6 ¹, Windows 2000¹, Windows XP¹)

PC requirements: Intel Pentium III, min. 1GHz, min. 256 MB RAM, minimum SuperVGA; 19" monitor recommended

Scanner: Fujitsu, and other TWAIN compatible scanner

¹⁾Registered trademark of Microsoft[™]
²⁾Registered trademark of Scansoft Inc
³⁾Registered trademark of Adobe Systems Inc.



Producer: ARPA Data GmbH

Albisstrasse 36 CH 8134 Adliswil Tel: +41 1 709 09 79

www.arpa.ch